Research Paper ■

Seeking Health Information Online: Does Wikipedia Matter?

MICHAËL R. LAURENT, TIM J. VICKERS

Abstract **Objective:** To determine the significance of the English Wikipedia as a source of online health information.

Design: The authors measured Wikipedia's ranking on general Internet search engines by entering keywords from MedlinePlus, NHS Direct Online, and the National Organization of Rare Diseases as queries into search engine optimization software. We assessed whether article quality influenced this ranking. The authors tested whether traffic to Wikipedia coincided with epidemiological trends and news of emerging health concerns, and how it compares to MedlinePlus.

Measurements: Cumulative incidence and average position of Wikipedia® compared to other Web sites among the first 20 results on general Internet search engines (Google®, Google UK®, Yahoo®, and MSN®), and page view statistics for selected Wikipedia articles and MedlinePlus pages.

Results: Wikipedia ranked among the first ten results in 71–85% of search engines and keywords tested. Wikipedia surpassed MedlinePlus and NHS Direct Online (except for queries from the latter on Google UK), and ranked higher with quality articles. Wikipedia ranked highest for rare diseases, although its incidence in several categories decreased. Page views increased parallel to the occurrence of 20 seasonal disorders and news of three emerging health concerns. Wikipedia articles were viewed more often than MedlinePlus Topic ($p = 0.001$) but for MedlinePlus Encyclopedia pages, the trend was not significant ($p = 0.07–0.10$).

Conclusions: Based on its search engine ranking and page view statistics, the English Wikipedia is a prominent source of online health information compared to the other online health information providers studied.

■ **J Am Med Inform Assoc.** 2009;16:471–479. DOI 10.1197/jamia.M3059.

Introduction

This paper evaluates the rate of occurrence of the English edition of Wikipedia, a large online collaborative encyclopedia, among the top results from leading general Internet search engines for health queries derived from three large online health information resources. We compared Wikipedia's occurrence and mean position to other Web sites, and examined which factors influence the Web site's position. We also investigated whether traffic to Wikipedia articles correlated with epidemiological factors, and how page views statistics compared to MedlinePlus, a major governmental online health encyclopedia.

Affiliations of the authors: Faculty of Medicine, Katholieke Universiteit Leuven (MRL), Belgium; Department of Molecular Microbiology, Washington University, School of Medicine (TJV), St Louis, MO, United States.

The authors thank S.C. Grover, MD, FRCPC and J.F. De Wolff, MD, MRCP for critically reviewing our manuscript.

This manuscript has not been submitted to any institutional review boards and has not received any funding or support.

The authors contribute as volunteers to the English Wikipedia (where both are administrators) and other Web sites of the Wikimedia Foundation. M. Laurent is a member of the WikiProject Medicine, and T. Vickers is Director of the Molecular and Cellular Biology WikiProject.

Correspondence: Michaël R. Laurent, Rooistraat 8, B-3012 Wilsele, Belgium; e-mail: <michael.laurent@gmail.com>.

Received for review: 11/04/08; accepted for publication: 04/01/09.

Background

Although the Internet is a copious source of health information, how and how often Internet users look for health information online and what the impact of this behavior has on the patient–physician relationship, remains unclear.[1–7] Online health information lookup is more frequent in certain populations, such as men and those with higher levels of education and health literacy, and less frequent in demographic groups such as elderly people.[1,8,9] Despite the existence of search engines designed to retrieve information from Web sites that have been assigned quality labels (such as the one assigned by the Health on the Net Foundation), general search engines (of which Google is the market leader in many Western countries) appear to be the most popular starting points for online health information searches.[4,5,8,10,11] Importantly, the first page of general search engine results is significantly more likely to be accessed by (inexperienced) health information seekers, with an exponential decline thereafter.[10,11] The rank of a Web site among search engine results depends on factors such as the specific search engine algorithm, the number of times the Web site is accessed from the results page (by the demographic group that uses the search engine), and search engine optimization strategies that aim to influence ranking.[12]

Wikipedia (http://www.wikipedia.org) is an open-access multilingual online encyclopedia that invites contributions from its users. It is operated as a charity by the non-profit Wikimedia Foundation. Wikipedia ranks as the eighth most accessed Web site on the Internet, according to Internet traffic information from Alexa, Inc.[13] With now more than 2.5 million articles, the English Wikipedia is the most prominent example of a wiki website. Wikis use a relatively simple editing syntax and a public record of all edits to facilitate collaboration between multiple contributors. While not a specific medical Internet encyclopedia like MedlinePlus[14] or NHS Direct Online,[15] Wikipedia contains articles on many medical topics.[16] In 2006, a small study using eight commercially obtained popular health-related search terms on Google and Yahoo (total of 16 searches) found that user-generated content appeared on the first page of results in 12 cases, ten of which included results from Wikipedia. For the search terms "diabetes" and "bipolar disorder", search engine statistics revealed that consumers visited Wikipedia in 2.79 and 7.17% of cases, respectively.[17]

Research Question

The frequency with which Wikipedia and many other health information resources feature on the first pages of search engine results has yet to be accurately determined. We believe such rankings are important because they provide insights into which Web sites are likely to be visited by online health information seekers. The aim of our study was to determine how often the English Wikipedia appears among the top search engine results for health-related queries (with the average rank as a secondary outcome), and how this compares to the position of governmental, non-governmental and commercial health information Web sites. Furthermore, we evaluated factors involved in Wikipedia's position by determining whether the quality of Wikipedia articles, as rated by its contributors, influenced this position (to test whether articles rated as more developed were ranked higher), and compared how Wikipedia ranked among search engine results when rare diseases were used as keywords, versus keywords containing more common health terms.

In addition, we investigated traffic trends for articles on medical conditions or pathogens that show seasonal variation. Our hypothesis was that if consumers use general search engines to seek health information online, they would commonly be exposed to Wikipedia content. If they would also access these search engine results, then traffic to articles with a season-specific topic would be predicted to increase parallel to endemic occurrence of the condition or its pathogen. In addition, news of a sudden infectious disease outbreak or other emerging health concern should cause a sudden increase in traffic to relevant articles. Finally, we compared page view statistics of MedlinePlus Topic and Encyclopedia pages to page views of the corresponding Wikipedia articles.

Methods

Software and Search Engines

We used a search engine optimization tool (Advanced Web Ranking, version 6.2, by Caphyon, Ltd, Craiova, Romania, available from http://www.advancedwebranking.com/index.html) to check the position of the English Wikipedia versus other domains among the first 20 results retrieved from Google (http://www.google.com), Google UK (http://www.google.co.uk), Yahoo (http://www.yahoo.com) and MSN (http://www.msn.com). The software determines the number of times the entered domains were found as the first result or among the first five, ten or twenty results. We entered these absolute cumulative incidences into a spreadsheet application (OpenOffice.org Calc version 2.4.0 by the OpenOffice.org community and Sun Microsystems, Inc, Santa Clara, CA, United States) to calculate relative cumulative incidences. We used the search engine optimization software to determine the position of a given domain among search engine results; we used these data to test whether the mean position was higher for community-rated quality articles on the English Wikipedia (*see below*, section "Influence of community-rated article quality").

Keywords

First, we extracted 1726 keywords from the health topic index of MedlinePlus, a health information service from the United States Library of Medicine and the National Institutes of Health (keywords listed at http://www.nlm.nih.gov/medlineplus/all_healthtopics.html; we used the Aug 12, 2008 update). The keywords included the titles of all MedlinePlus Health Encyclopedia entries, as well as common synonyms, related search terms and abbreviations (for example, "Attention Deficit Hyperactivity Disorder", "ADHD", "ADD" and "Hyperactivity" were all included). Some of the keywords (such as "Alaska Native Health") are specific to the United States in context, and the spelling of keyword terms is in American English. When search terms consisted of exactly the same words (e.g., "Allergy, Food" and "Food Allergy"), we removed one version from the list. A few obviously irrelevant keywords like "Teens page" were removed as well, but no other terms were removed (for example, "Terrorist Attacks" and "Tornadoes" were not rejected as keywords).

We then derived a second set of all 966 keywords in the online alphabetical topic index of NHS Direct Online, a health information service from the British National Health Service (keywords listed at http://www.nhsdirect.nhs.uk/encyclopaedia/a-z/, retrieved on Aug 22, 2008). These keywords were in British English and included terms specifically used in the UK (e.g., "A&E" and "NHS"). Both sets of keywords represent a large number of common medical conditions, symptoms, diagnostic tests, anatomical and physiological terms, treatments, procedures, prevention topics and other medical terms.

Finally, we derived 1,173 keywords from the online alphabetical index of the U.S. National Organization of Rare Diseases (NORD; keywords listed at http://www.rarediseases.org/search/rdblist.html, retrieved on Sept 12, 2008). This list unexpectedly contained keywords referring to conditions that are not rare, such as cataracts, carpal tunnel syndrome, colon cancer, and prostate cancer. We used the software tool to check the position of selected Web sites for the first two sets of keywords between Aug 19 and 23, 2008, and for the third set of keywords on Sept 12 and 13, 2008 (the full list of keywords is listed in Appendix 1, available as an online data supplement at http://www.jamia.org).

Web Sites and Domains

We selected 27 Web sites or groups of Web sites for comparison to Wikipedia based on their ranking on manual searches during the preparation of this paper. This involved using keywords from MedlinePlus as queries on Google and including those content providers that frequently appeared among the first 20 results, as based on manual ranking calculations (data not shown). Some Web sites were included because they belong to notable organizations (e.g., the World Health Organization or the Health on the Net Foundation), even though they frequently ranked lower than some other commercial Web sites that were not included because of a narrow scope (e.g., http://www.drugs.com or http://www.babycenter.com).

The software tool we used did not recognize subdomains (e.g., the results for http://nlm.nih.gov were not shown under http://nih.gov). Accordingly, after selecting the online health information providers, we grouped Internet domains belonging to a single organization so that they would be processed together. We manually searched for additional subdomains to ensure maximal clustering. We created a "U.S. government" cluster which consisted of 123 ".gov" Internet domains (ranging from http://www.4woman.gov to http://www.womenshealth.gov). These domains were identified among Google's first 200 results for "site:*.gov health", and additional listing of all domains of United States National Institutes of Health centers and institutes. We created an "http://about.com" cluster with 96 subdomains (ranging from http://acne.about.com to http://yoga.about.com) based on the list of available subdomains on the provider's Web site. As well as analyzing these Web sites separately, we grouped together commercial Web sites belonging to WebMD (http://www.webmd.com, http://www.emedicine.com, http://www.emedicinehealth.com and http://www.medscape.com). Furthermore, we paired http://www.familydoctor.org with http://www.aafp.org, as both are maintained by the American Academy of Family Physicians. Finally, we clustered three domains related to the British Broadcasting Company (BBC), four domains and subdomains related to the Cleveland Clinic and two related to the Mayo Clinic (all domain clusters are listed in Appendix 2, available as an online data supplement at http://www.jamia.org). The software tool automatically allocates the ranking from the highest listed domain to the cluster to which it belongs.

Influence of Community-rated Article Quality

The English Wikipedia allows groups of editors collaborating in a certain area of knowledge (called a WikiProject[18,19]) to assess the quality of articles in their field. Of the possible quality ratings (which, from lowest to highest rating, are termed Stub, Start, C-class, B-class, Good Article, A-class and Featured Article), only two are applied after a formal review process: Featured Article (which are meant to exemplify Wikipedia's best work) and Good Articles (which are judged by similar but less stringent criteria). We identified all health-related Featured and Good Articles (hereafter referred to as "quality articles") via their respective categories and index pages (see http://en.wikipedia.org/wiki/Category:Medicine_articles_by_quality for an overview). For these quality articles, we looked for equivalent MedlinePlus keywords; 49 out of 1726 keywords (2.8%) had corresponding quality articles on the English Wikipedia. Using the search engine optimization tool described above, we tested whether these quality articles were listed more frequently among the first 20 search engine results for MedlinePlus keywords, and if they had higher mean positions compared to non-quality Wikipedia articles.

Page Views of Season-related Articles and Emerging Health Concerns

We selected ten Wikipedia articles on conditions more common in winter, or on pathogens causing an illness that is more common during winter: frostbite, hypothermia, carbon monoxide poisoning, common cold, pneumonia, bronchiolitis, norovirus, influenza, rhinovirus and seasonal affective disorder. In the same manner, we selected articles related to the summer: hyperthermia, sunburn, hay fever, insect bites and stings, bee sting, Lyme disease, Rocky Mountain spotted fever, hemolytic-uremic syndrome, harvest mite and West Nile virus. We retrieved available information on daily page views for these articles from http://stats.grok.se, and compared daily page views from Jun to Jan 2008. To complement this study of seasonal epidemiological influences, we also studied whether emerging health concerns influenced the number of page views of the relevant Wikipedia article. Here, we studied three examples for which the United States Centers for Disease Control and Prevention and/or the Food and Drug Administration issued alerts. The first concerns melamine-contaminated infant formula from China, which first received broad media attention on Sept 12, 2008. The second example involves the *Salmonella* Saintpaul outbreak, which led several groceries and restaurants to stop offering tomatoes on Jun 9, 2008. The third example involves an intoxication with the protein toxin ricin which was announced on Feb 29, 2008. We therefore retrieved page view statistics for the Wikipedia articles "Melamine", "Salmonella", and "Ricin".

Traffic to Medlineplus Compared to Wikipedia

We obtained page view statistics for the 20 most visited MedlinePlus Topic and Encyclopedia pages, both for Jan and Jun 2008 (personal communication, Kitendaugh P, Head of Reference and Web Services, Public Services Division of the U.S. National Library of Medicine). We then compared the MedlinePlus page view statistics to those obtained from http://stats.grok.se for the corresponding Wikipedia article (if one existed). To determine the corresponding Wikipedia article, we entered the MedlinePlus term into Wikipedia's search box and used the first applicable result.

Statistical Analyses

Statistical analyses were performed using the GraphPad Prism software (version 5.01, by GraphPad Software, Inc, La Jolla, CA, United States). We defined $P < 0.05$ as statistically significant. Differences between proportions were determined using contingency tables and the χ^2 test (or Fisher's exact test for the smaller sample of quality articles). Differences between means were determined using two-sided Student's t tests. To assess whether news of emerging health concerns influenced traffic to Wikipedia articles, we used a one-sample t test to determine whether the mean number of daily page views was different from the highest number during that month. Page views of MedlinePlus and corresponding Wikipedia articles were compared using a paired t test.

Table 1 ■ Cumulative Incidence of Selected Domains for Queries on Google

	MedlinePlus Queries, No. (%)				
Domains	First Place	Top 5	Top 10	Top 20	Domains
English Wikipedia	**585 (33.9)**	**1173 (68.0)**	**1285 (74.5)**	1352 (78.3)	English Wikipedia
.gov domains*	420 (24.3)	1006 (58.3)	**1282 (74.3)**	**1407 (81.5)**	NHS Direct Online
MedlinePlus	289 (16.5)	719 (41.7)	1059 (61.4)	1192 (69.1)	Medscape*
Medscape*		390 (22.6)	749 (43.4)	971 (56.3)	.gov domains*
Mayo Clinic*		359 (20.8)	546 (31.6)	683 (39.6)	MedlinePlus
Medicinenet.com		270 (15.6)	475 (27.5)	654 (37.9)	Mayo Clinic*
eMedicine.com		166 (9.62)	391 (22.7)	558 (32.3)	Medicinenet.com
KidsHealth.org		176 (10.2)	285 (16.5)	362 (21.0)	WebMD.com
WebMD.com		103 (5.97)	270 (15.6)	473 (27.4)	eMedicine.com
Emedicinehealth.com		142 (8.23)	258 (15.0)	359 (20.8)	Emedicinehealth.com
AAFP*			167 (9.68)	279 (16.2)	KidsHealth.org
About.com*			162 (9.39)	320 (18.5)	Patient.co.uk
Merck.com			160 (9.27)	302 (17.5)	NetDoctor.co.uk
NHS Direct Online			136 (7.88)	323 (18.7)	BBC*
Patient.co.uk			101 (5.85)	269 (15.6)	AAFP*
NetDoctor.co.uk			93 (5.39)	194 (11.2)	Merck.com
WrongDiagnosis.com			91 (5.27)	154 (8.92)	About.com*

AAFP = American Academy of Family Physicians; NHS = National Health Service; NORD = National Organization of Rare Diseases.
This table shows cumulative incidences of selected domains for queries from MedlinePlus, NHS Direct Online and the National Organization of Rare Diseases (NORD) on Google.
Bold text denotes that these numbers are significantly ($p < 0.05$) larger than the numbers below in the same column. Similarly, underlined text denotes significantly lower measurements for NORD keywords compared to MedlinePlus keywords. Only numbers with percentages ≥ 5% are shown, and domains not reaching a cumulative incidence ≥ 5% among the first ten results are not tabulated (for the full results on all search engines and domains, see Tables 3–6, available as an online data supplement at http://www.jamia.org). Domains are sorted in descending order according to Top 10 incidence for MedlinePlus queries.
*=These are domain clusters, see "Methods" under "Websites" for details. AAFP.

Results

Incidence and Mean Position of Studied Domains Among Search Engine Results

Table 1 shows the domains with the highest cumulative incidences for queries on Google. Among Google's results for queries from MedlinePlus, the English Wikipedia ranked highest among the first results and the first five results. The United States government cluster tied with the English Wikipedia among the first ten results, and surpassed it among the first 20 results. Google more commonly listed results from the English Wikipedia than from MedlinePlus or NHS Direct Online, even when they were the source of the search terms. When using only rare diseases as keywords, the English Wikipedia outranked all other sites, but its incidence showed only small and inconsistent differences compared to its results for MedlinePlus queries. On queries

Table 2 ■ Cumulative Incidence of Selected Domains for Queries on Google U.K.

	NHS Direct Online Queries, No. (%)					MedlinePlus Queries, No. (%)			
Domains	First Place	Top 5	Top 10	Top 20	Domains	First Place	Top 5	Top 10	Top 20
NHS Direct Online	212 (22.0)	**660 (68.3)**	**800 (82.8)**	**848 (87.8)**	English Wikipedia	**640 (37.1)**	**1096 (63.5)**	**1222 (70.8)**	**1300 (75.3)**
English Wikipedia	**293 (30.3)**	567 (58.7)	657 (68.0)	708 (73.3)	.gov domains*	254 (14.7)	493 (28.6)	681 (39.5)	854 (49.5)
BBC*		254 (26.3)	474 (49.1)	585 (60.6)	BBC*		312 (18.1)	610 (35.3)	775 (44.9)
Patient.co.uk		246 (25.5)	427 (44.2)	557 (57.7)	MedlinePlus	210 (12.2)	372 (21.6)	522 (30.2)	652 (37.8)
NetDoctor.co.uk		190 (19.7)	340 (35.4)	405 (41.9)	Patient.co.uk		256 (14.8)	480 (27.8)	645 (37.4)
Medscape*			95 (9.8)	161 (16.7)	NHS Direct Online		317 (18.4)	425 (24.6)	473 (27.4)
.gov domains*			73 (7.6)	128 (13.3)	NetDoctor.co.uk		187 (10.8)	346 (20.1)	407 (23.6)
MedlinePlus			73 (7.6)	126 (13.0)	Medscape*		151 (8.75)	282 (16.3)	412 (23.9)
eMedicine.com			56 (5.8)	89 (9.2)	eMedicine.com		105 (6.08)	186 (10.8)	265 (15.4)
					Medicinenet.com			147 (8.52)	261 (15.1)
					KidsHealth.org			102 (5.91)	149 (8.63)
					Mayo Clinic*			97 (5.6)	184 (10.7)

NHS = National Health Service; BBC = British Broadcasting Company.
This table shows cumulative incidences of selected domains for NHS Direct Online and MedlinePlus queries on Google UK. Bold text denotes that these numbers are significantly ($P < 0.05$) larger than the numbers below in the same column. Only numbers with percentages ≥ 5% are shown; domains not reaching this threshold for the first ten results on Google UK are not tabulated (for the full results on all search engines and domains, see Tables 3–6, available as an online data supplement at http://www.jamia.org). Domains are sorted in descending order according to Top 10 incidence for MedlinePlus queries.
*=These are domain clusters, see "Methods" under "Websites" for details. BBC = British Broadcasting Company.

Table 1 ■ (continued)

NHS Direct Online Queries, No. (%)					NORD Queries, No. (%)			
First Place	Top 5	Top 10	Top 20	Domains	First Place	Top 5	Top 10	Top 20
345 (35.7)	**684 (70.8)**	**749 (77.5)**	**785 (81.3)**	English Wikipedia	**289 (24.6)**	**740 (63.1)**	**830 (70.8)**	**889 (75.8)**
62 (6.4)	290 (30.0)	494 (51.1)	719 (74.4)	Medscape*	156 (13.3)	502 (42.8)	691 (58.9)	789 (67.3)
	248 (25.7)	444 (46.0)	573 (59.3)	eMedicine.com	129 (11.0)	401 (34.2)	545 (46.5)	640 (54.6)
	208 (21.5)	421 (43.6)	519 (53.7)	.gov domains*	102 (8.70)	335 (28.6)	469 (40.0)	578 (49.3)
	205 (21.2)	417 (43.2)	516 (53.4)	WrongDiagnosis.com		195 (16.6)	395 (33.7)	585 (49.9)
	199 (20.6)	328 (34.0)	420 (43.5)	MedlinePlus		158 (13.5)	321 (27.4)	371 (31.6)
	188 (19.5)	322 (33.3)	420 (43.5)	Mayo Clinic*		134 (11.4)	209 (17.8)	247 (21.1)
	87 (9.0)	202 (20.9)	303 (31.4)	WebMD.com		108 (9.21)	205 (17.5)	299 (25.5)
	93 (9.6)	199 (20.6)	302 (31.3)	Medicinenet.com		87 (7.4)	160 (13.6)	201 (17.1)
	89 (9.2)	165 (17.1)	242 (25.1)	About.com*		61 (5.2)	147 (12.5)	216 (18.4)
	101 (10.5)	165 (17.1)	212 (22.0)	Merck.com			121 (10.3)	189 (16.1)
	81 (8.4)	154 (16.0)	295 (30.5)	Patient.co.uk			100 (8.53)	228 (19.4)
	83 (8.6)	141 (14.6)	246 (25.5)	Medterms.com			87 (7.4)	166 (14.2)
	57 (5.9)	126 (13.0)	299 (30.6)	KidsHealth.org			61 (5.2)	72 (6.1)
		99 (10)	163 (16.9)					
		74 (7.7)	158 (16.4)					
		69 (7.1)	162 (16.8)					

from NHS Direct Online and NORD, the Medscape cluster had higher incidences than the United States government cluster. Table 2 allows comparison of results for NHS Direct Online and MedlinePlus queries on Google U.K. Except among first results, NHS Direct Online was listed more frequently when its keywords were used. The BBC ranked third on Google U.K. Overall, the English Wikipedia ranked among the first ten results in between 70.8 and 84.7% of cases across search engines and keywords. For more detailed incidence statistics on all Web sites and search engines, see Tables 3–6, available as online data supplements at http://www.jamia.org. For MedlinePlus keywords, the English Wikipedia's mean ranking across search engines was the third position (2.77–3.58 across search engines), which was higher than other domains (see Table 7, available as an online data supplement at http://www.jamia.org).

Influence of Community-rated Article Quality

When the MedlinePlus keywords related to the 49 quality Wikipedia articles were used as queries, the Wikipedia articles were found among the first ten results in all cases on Google and Google UK, and in 47 cases (96%) on Yahoo and MSN. This was significantly more frequently than the top ten and top 20 incidences of Wikipedia for all MedlinePlus keywords (compared to the top 20 incidences of 86% on Yahoo and 85% on MSN, $p = 0.04$ and $p = 0.02$, respectively).

Epidemiological Influences on Wikipedia Article Page Views

Figure 1 shows the relative amount of page views for Jun compared to the mean number of daily page views in Jan for ten conditions or pathogens that occur more commonly during either winter or summer months. All these articles had significant differences in daily traffic between these two months (t test for all $p < 0.0001$ except for hypothermia, $p = 0.0002$). Figure 2 shows the daily page views of three Wikipedia articles related to emerging health threats. These increases could not be attributed to a normal variance (one-sample t test of highest value compared to mean over days before incident: all three $p < 0.0001$).

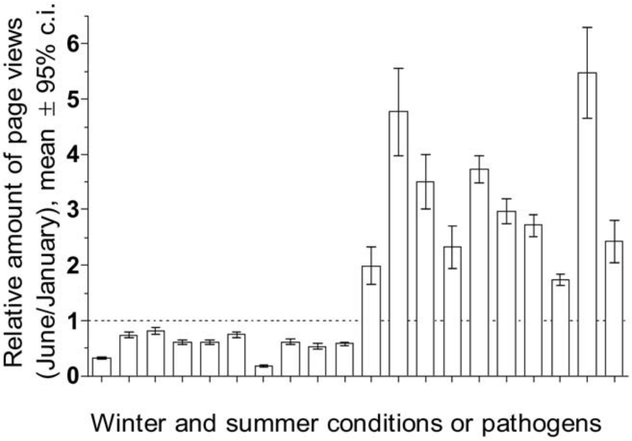

Figure 1. This graph shows the amount of page views during Jun relative to Jan 2008 for ten conditions or pathogens which are more common during winter (shown left: frostbite, hypothermia, carbon monoxide poisoning, common cold, pneumonia, bronchiolitis, norovirus, influenza, rhinovirus and seasonal affective disorder) or summer months (shown right: hyperthermia, sunburn, hay fever, insect bites and stings, bee sting, Lyme disease, Rocky Mountain spotted fever, hemolytic-uremic syndrome, harvest mite, West Nile virus), respectively. Results are expressed as a mean relative number of page views, and bars symbolize the 95% confidence interval around the mean.

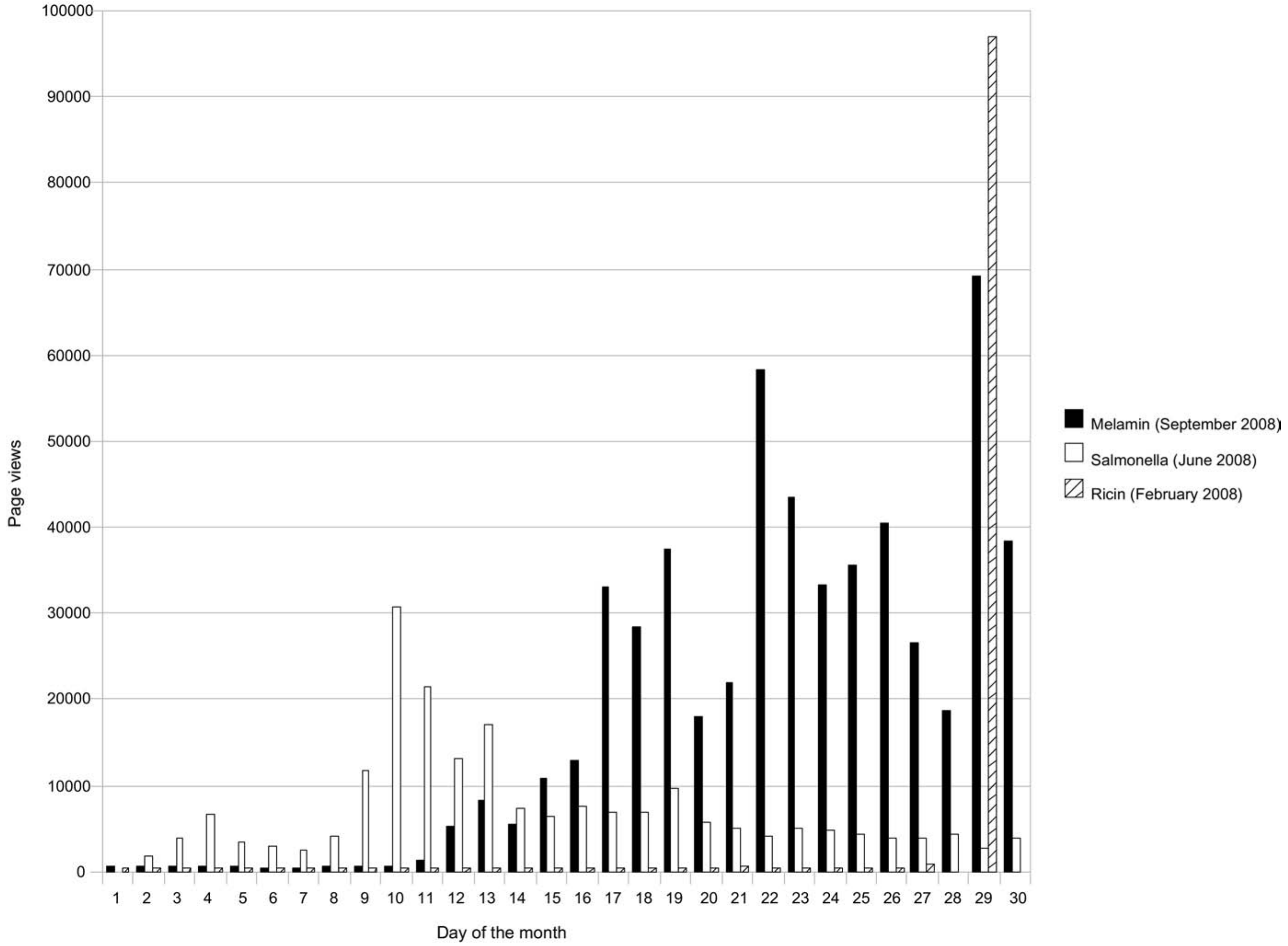

Figure 2. This graph shows the daily amount of page views during Sept 2008 for the article "Melamine" on the English Wikipedia, the Jun 2008 daily page views for "Salmonella", and the Feb 2008 daily page views for the article "Ricin". On Sept 12, 2008, reports emerged of melamine-contaminated infant formula in China. Following the announcement of an outbreak of *Salmonella* Saint Paul and a possible link to tomatoes, several groceries and restaurants stopped offering tomatoes on Jun 9, 2008. On Feb 29, 2008, the Centers for Disease Control and Prevention announced an investigation following an intoxication with the protein toxin ricin.

Page Views of MedlinePlus Versus Wikipedia

Wikipedia's articles were viewed more frequently than the corresponding MedlinePlus Topic pages ($p = \text{or} < 0.001$); there was a non-significant trend towards higher page views for Wikipedia compared to MedlinePlus Encyclopedia pages ($p = 0.068$ and $p = 0.097$ for Jan and Jun 2008, respectively). Complete page view statistics are given in Table 8, available as an online data supplement at http://www.jamia.org.

Discussion

The aim of this paper was to determine the relative position of the English Wikipedia and other Web sites containing health information in a search engine-based approach. The results show that if the first page of results of a general search engine lists ten Web sites, Wikipedia can be found among those results in more than 70% of cases. This confirms preliminary findings by others.[17] Wikipedia had a higher average position than any other reference in this study. Our findings on resources other than Wikipedia confirm previous findings using Internet audience measurement services, which did not include Wikipedia's medical content.[20]

Wikipedia ranked higher with quality articles, although this is not necessarily a causal relationship since these quality articles covered more common health topics, and we have observed that Wikipedia was more prominent among search results for common health terms in some categories. Wikipedia's good results for rare diseases compared to other online health resources also suggest that it has articles on a wide range of conditions. The results pertaining short- and long-term epidemiological influences on article traffic create a link between search engine results and page viewing. Others have previously observed the relationship between search engine activity and news coverage.[21] A study on Google Flu showed that search engine queries related to influenza-like illness correlated with the epidemiological data from the United States Centers for Disease Control and Prevention.[22] These findings were replicated for queries submitted to a Swedish medical Web site.[23] We believe that these studies support our assumptions that firstly, Internet activity can be used as a surrogate marker for consumer behavior, and secondly, that online health information seekers often use search engines to find individual health Web sites, which underscore the importance of Wikipedia as a prominent source of information in such searches.

Our study has several strengths. The software tool we used allowed us to check a large set of keywords on multiple search engines while avoiding observer bias. The use of a broad set of keywords from governmental online health information initiatives was important to avoid selection bias, and additionally it might make these data useful from a policy-making point of view; we provide data that might be relevant to the search engines position of the government-sponsored health information Web sites MedlinePlus and NHS Direct Online.

However, our study design does have limitations. We did not perform a weighted analysis based on often-used health-related keywords (such as "Diabetes");[3,17] so in our study, each keyword was given equal importance. Some keywords were listed together with their abbreviations, which results in multiple counting. Nevertheless, this could mimic how people use search engines, with some people using an abbreviation while others might know the full term. Indeed, consumers appear to differ widely in the queries they use to find specific information,[24] which also limits the generalizability of our search terms. Because we wanted to avoid selection biases, we also retained keywords that returned several non-medical Web sites as search engine results (for words like "Walkers" or abbreviations like "CFS" for chronic fatigue syndrome), which favors Wikipedia because it contains more than just medical information. Unexpectedly, the conditions listed by the NORD contained some fairly common disorders (see Methods); we did not remove these, again to avoid selection bias. We made a personal selection of commercial, non-profit and governmental Web sites based on manual searches on Google for comparison to Wikipedia, but our list is by no means exhaustive and has the serious drawback of possible selection bias. However, the software allows storage of the data set and post hoc analysis for additional domains. We also created several clusters to pool the impact of a single content provider who might use multiple domains; we cannot completely exclude that some less prominent United States government Web sites were not listed and might not have been counted, although we believe that any such Web sites were unlikely to have a major impact on the results. It should also be noted that we did not study sponsored search engines results, which might influence consumers. We have tried to make a simple dichotomy (quality vs. non-quality articles) based on Wikipedia's community article rating system, but we emphasize that this is a system that has not yet been externally validated as a true measure of quality (compared to expert review). The examples we have provided of real-life epidemiological changes correlating with page views of the relevant Wikipedia article are illustrative, although this remains indirect evidence. There may be other seasonal disorders or pathogens that do not follow this pattern, and disease outbreaks that do not result in increased article traffic. The latter examples may be confounded by Wikipedia's role as a source of news, as disease outbreaks straddle the border between health information and news.

With regards to the generalizability of these results, it should again be stressed that not all online health information seekers are patients, and that not all patients seek health information online. Obviously, this study says little about consumers with a native language different from English[25] or using search engines popular in other countries (like http://Baidu.com in the People's Republic of China and http://Guruji.com in India). In this aspect, the results from British versus American English keywords are not mutually exchangeable. Furthermore, this Internet study provides no evidence on the level of trust that patients assign to the health information they read on Wikipedia. However, a recent survey indicated a shift of priorities for both non-professional and professional Internet users from trustworthiness and accuracy of information to availability and ease of finding information.[5] Finally, differences could exist between medical specialties with regards to the importance of both general health information Web sites and Web sites devoted to a specific topic.

Although several medical scientists and policy makers have highlighted the potential use of wikis to foster collaboration on easily-accessible health information for the community,[26–30] and Wikipedia is the most prominent example of a wiki, we could identify no previous research specifically focusing on Wikipedia as a source of health information for consumers. Thus, it appears that Wikipedia provides an important area for future research on sources of online health information. However, we also found misconceptions about Wikipedia in the scientific literature: for example, a recent study examining search engine results for obstetric queries misclassified it as a commercial instead of a non-profit Web site.[31] Importantly, the open editing policy offers no way of assessing the expertise of contributors, resulting in fears of inaccuracies. This may be one reason why doctors are creating wikis where only they can contribute (such as http://Ganfyd.org, http://RadiologyWiki.org or http://WikiSurgery.com).[32–34] Instead of creating new wikis, Wikipedia itself could be used by doctors, as well as patient groups and associations, to collaboratively edit articles on the topics they value.[17,35] As we have shown here, these articles are among the top results on general search engines, thus providing a free platform to disseminate information globally. Until now, doctors have lagged behind biomedical scientists in realizing the potentials of wikis.[18,36–51] However, in Feb 2009, Medpedia, Inc, in collaboration with several prominent medical faculties, the NHS, the American College of Physicians and other partners, launched its open access medical encyclopedia running on the same software and under the same license as Wikipedia. It will have content for the public as well as for experts, and allow for discussion of the subject. Contrary to Wikipedia where anyone can contribute regardless of qualifications, only experts are allowed to contribute to Medpedia (although others may suggest changes), which might alleviate quality concerns. We believe this wiki may address some of the concerns that discourage the medical community from contributing to Wikipedia.

Although this Internet study showed that Internet consumers are likely to be exposed to Wikipedia through search engine results for health-related keywords, examining quality of health information present in Wikipedia was beyond the scope of this article. Thus, further studies are urgently needed to determine whether Wikipedia articles are of sufficient quality to support patient-provider communication. Assessment of the quality of Wikipedia's freely editable content is difficult, since its articles are inherently in a constant state of flux. Examples of flagrant mistakes have been reported in the media, as well as an analysis that found that Wikipedia contains a similar numbers of mistakes compared to Encyclopedia Brittannica.[52] Clauson et al (2008) compared drug information from Wikipedia to the Medscape Drug Reference, and concluded that although Wikipedia was less complete (especially regarding dosing information, which is explicitly discouraged by Wikipedia guidelines), no factual errors were found, and it "may be a useful point of engagement for consumers" for supplemental drug information.[53] Although articles in the English Wikipedia are increasingly being referenced with articles from leading scientific journals,[16,54] and articles have been shown to improve over time,[53] Wikipedia itself makes no claim to correctness, and the medical disclaimer aptly describes the situation: "Wikipedia contains articles on many medical topics; however, no warranty whatsoever is made that any of the articles are accurate." However, while consumers appear to rarely check the source and quality of the information they find online;[4,9] at least with a well-known brand like Wikipedia, they know that they should remain skeptical. Indeed, while Wikipedia contributors have classified over 14,000 of their articles as dealing with medical topics: only around 50 of them have been confirmed as top quality ("Featured articles").[55] This implies that Wikipedia is still a long way from achieving the idea of its founder Jimmy Wales, who imagined a world where every human being had free access to the sum of all human knowledge in his or her language.[56] Maybe doctors, like researchers, "should read Wikipedia cautiously and amend it enthusiastically",[57] thus fulfilling the proposed new professional obligation of making their knowledge and expertise freely available on the Internet.[58]

Conclusions

Our study shows that Wikipedia is a prominent health information Web site based on its position among search engine results for health-related queries. Despite several calls to adopt the principles that underlie its success, there is virtually no research on Wikipedia's role as a source of health information. Observational studies in different settings are needed to document how often consumers seek health information online, and studies in a clinical setting are needed to estimate the impact of this behavior on the patient–physician relationship.

References ▪

1. Baker L, Wagner TH, Singer S, Bundorf MK. Use of the internet and e-mail for health care information: Results from a national survey. J Am Med Assoc 2003;289(18):2400–6.
2. Eysenbach G, Köhler C. Health-related searches on the internet. J Am Med Assoc 2004;291(24):2946.
3. Phillipov G, Phillips PJ. Frequency of health-related search terms on the internet. J Am Med Assoc 2003;290(17):2258–9.
4. Fox S. Online health Search 2006. In: Pew Internet and American Life Project; Washington. vol DC, 2006.
5. Health on the Net Foundation. Analysis of 9th HON survey of health and medical internet users, winter 2004–2005 Available at: http://www.hon.ch/Survey/Survey2005/res.html. Accessed November 2008.
6. Harris Interactive. Harris Poll shows Number of "Cyberchondriacs"—Adults who have ever gone online for health information—Increases to an estimated 160 million nationwide Available at: http://www.harrisinteractive.com/harris_poll/index.asp?PID=792. Accessed November 2008.
7. Gerber B, Eiser A. The patient–physician relationship in the internet age: Future prospects and the research agenda. J Med Internet Res 2001;3(2):e15.
8. Dickerson S, Reinhart AM, Feeley TH, et al. Patient internet use for health information at three urban primary care clinics. J Am Med Inform Assoc 2004;11(6):499–504.
9. National Center for Education Statistics. The health literacy of America's adults: Results from the 2003 national assessment of adult literacy Available at: http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2006483. Accessed November 2008.
10. Eysenbach G, Köhler C. How do consumers search for and appraise health information on the World Wide Web? Qualitative study using focus groups, usability tests, and in-depth interviews. BMJ 2002;324(7337):573–7.
11. Hansen DL, Derry HA, Resnick PJ, Richardson CR. Adolescents searching for health information on the internet: An observational study. J Med Internet Res 2003;5(4):e25.
12. Greenberg L, D'Andrea G, Lorence D. Setting the public agenda for online health search: A white paper and action agenda. J Med Internet Res 2004;6(2):e18.
13. Alexa. Top 500 sites Available at: http://www.alexa.com/site/ds/top_sites?ts_mode=global. Accessed November 2008.
14. Miller N, Lacroix EM, Backus JE, MedlinePlus. Building and maintaining the National Library of Medicine's consumer health Web Service. Bull Med Libr Assoc 2000;88(1):11–7.
15. Gann B. NHS direct online: A multi-channel eHealth service. Stud Health Technol Inform 2004;100:164–8.
16. Czarnecka-Kujawa K, Abdalian R, Grover SC. The quality of open access and open source internet material in gastroenterology: Is Wikipedia appropriate for knowledge transfer to patients? Gastroenterology 2008;134(4):A-325.
17. Envision Solutions, LLC. Diving Deeper Into Online Health Search - Examining Why People Trust Internet Content & The Impact Of User-Generated Media. Available at: http://www.envisionsolutionsnow.com/pdf/Studies/Online_Health_Search.pdf. Accessed November 2008.
18. Daub J, Gardner PP, Tate J, et al. The RNA WikiProject: Community annotation of RNA families. RNA 2008;14:2462–4.
19. Ward R. A request for help to improve the coverage of the NHS and UK healthcare issues on Wikipedia. Health Info Internet. 2006;53(1):7–8.
20. Wood FB, Benson D, LaCroix EM, Siegel ER, Fariss S. Use of internet audience measurement data to gauge market share for online health information services. J Med Internet Res 2005;7(3):e31.
21. Cooper CP, Mallon KP, Leadbetter S, Pollack LA, Peipins LA. Cancer internet search activity on a major search engine, United States 2001–2003. J Med Internet Res 2005;7(3):e36.
22. Ginsberg J, Mohebbi MH, Patel RS, et al. Detecting influenza epidemics using search engine query data. Nature 2009;457(7232):1012–4.
23. Hulth A, Rydevik G, Linde A. Web queries as a source for syndromic surveillance. PLoS ONE 2009;4(2):e4378.
24. Toms EG, Latter C. How consumers search for health information. Health Inform J 2007;13(3):223–35.
25. Singh PM, Wight CA, Sercinoglu O, et al. Language preferences on websites and in Google searches for human health and food information. J Med Internet Res 2007 Jun 28;9(2):e18.
26. Giustini D. How web 2.0 is changing medicine. BMJ 2006;333(7582):1283–4.

27. Potts HW. Is e-health progressing faster than e-health researchers? J Med Internet Res 2006;8(3):e24.
28. Crespo R. Virtual community health promotion. Prev Chronic Dis 2007;4(3):A75.
29. Navarro A, Voetsch K, Liburd L, Bezold C, Rhea M. Recommendations for future efforts in community health promotion—Report of the National Expert Panel on Community Health Promotion Available at: http://www.cdc.gov/NCCDPHP/pdf/community_health_promotion_expert_panel_report.pdf. Accessed November 2008.
30. Altmann U. Representation of medical informatics in the Wikipedia and its perspectives. Stud Health Technol Inform 2005;116:755–60.
31. Kaimal AJ, Cheng YW, Bryant AS, et al. Google obstetrics: Who is educating our patients? Am J Obstet Gynecol 2008;198(6)(682):e1–5.
32. Keim B. News feature: WikiMedia. Nat Med 2007;13(3):231–3.
33. Streeter JL, Lu MT, Rybicki FJ. Informatics in radiology: RadiologyWiki.org: The free radiology resource that anyone can edit. Radiographics 2007;27(4):1193–200.
34. Agha R. Introducing Wikisurgery.com: The blueprint for a surgical architecture of participation. Int J Surg 2006;4(3):140–3.
35. Yager K. Wiki Ware could harness the internet for science. Nature 2006;440(7082):278.
36. Hoffmann R. A wiki for the life sciences where authorship matters. Nat Genet 2008;40(9):1047–51.
37. Huss JW III, Orozco C, Goodale J, et al. A gene Wiki for community annotation of gene function. PLoS Biol 2008;6(7):e175.
38. Hodis E, Prilusky J, Martz E, et al. A scientific "wiki" bridging the rift between three-dimensional structure and function of biomacromolecules. Genome Biol 2008;9(8):R121.
39. Stokes TH, Torrance JT, Li H, Wang MD. ArrayWiki: an enabling technology for sharing public microarray data repositories and meta-analyses. BMC Bioinform 2008;9 (Suppl 6):S18.
40. Pawlicki S, Le Béchec A, Delamarche CA. Pdb: A database dedicated to amyloid precursor proteins. BMC Bioinform 2008;9:273.
41. Mons B, Ashburner M, Chichester C, et al. Calling on a million minds for community annotation in WikiProteins. Genome Biol 2008;9(5):R89.
42. Wang X. miRDB: A microRNA target prediction and functional annotation database with a Wiki interface. RNA 2008 Jun;14(6):1012–7.
43. Mahapatra A. Catalyzing chemical bonding—The WIKI way. ACS Chem Biol 2007;2(12):755–7.
44. Osborne JD, Lin S, Kibbe WA. Other riffs on cooperation are already showing how well a Wiki could work. Nature 2007;446(7138):856.
45. Salzberg SL. Genome re-annotation: A Wiki solution? Genome Biol 2007;8(1):102.
46. Pearson H. Online methods share insider tricks. Nature 2006;441(7094):678.
47. Csõsz E, Meskó B, Fésüs L. Transdab wiki: The interactive transglutaminase substrate database on web 2.0 surface. Amino Acids 2009;36:615–7.
48. Pico AR, Kelder T, van Iersel MP, et al. Pathway editing for the people. PLoS Biol 2008;6:e184.
49. Hu JC, Aramayo R, Bolser D, et al. The emerging world of Wikis. Science 2008;320:1289–90.
50. Waldrop M. Big data: Wikiomics. Nature 2008;455:22–5.
51. Bidartondo MI. Preserving accuracy in GenBank. Science 2008;319:1616.
52. Giles J. Internet encyclopaedias go head to head. Nature 2005;438(7070):900–1.
53. Clauson KA, Polen HH, Boulos MN, Dzenowagis JH. Scope, completeness, and accuracy of drug information in Wikipedia. Ann Pharmacother 2008;42(12):1814–21.
54. Nielsen FA. Scientific citations in Wikipedia. First Monday 2008;12(8).
55. English Wikipedia contributors. Wikipedia:WikiProject Medicine/Assesment. Available at: http://en.wikipedia.org/wiki/Wikipedia:WikiProject_Medicine/Assessment. Accessed November 2008.
56. Khamsi R. Reference revolution. Nature. Published online 2005 Mar 18.
57. Editorial: Wiki's wild world. Nature 2005;438(7070):890.
58. Midgley A. Global medical knowledge database. New professional obligation arises. BMJ 2000;321(7267):1020.